First Annual Report for the Land Information System

Submitted under Task Agreement GSFC-CT-2

Cooperative Agreement Notice (CAN) CAN-00OES-01

Increasing Interoperability and Performance of Grand Challenge Applications in the Earth, Space, Life, and Microgravity Sciences

Period Covered: January 16, 2002-August 30, 2002

Version 1.0

Table of Contents

List of Figures

Acronyms and Terms

ALMA: Assistance for Land-surface Modeling Activities

API: Application Programming Interface

CGI: Common Gateway Interface

CLM: Community Land Model

DODS: Distributed Ocean Data System

ESMF: Earth System Modeling Framework

GrADS: Grid Analysis and Display System

LDAS: Land Data Assimilation System

LIS: Land Information System

MRTG: Multi Router Traffic Grapher

NFS: Network File System

NOAH: National Centers for Environmental Prediction, Oregon State University, United States Air Force, and Office of Hydrology Land Surface Model

PXE: Preboot Execution Environment

RAID: Redundant Array of Inexpensive Disks

SNMP: Simple Network Management Protocol

VIC: Variable Infiltration Capacity Land Surface Model

1 Objective

This Annual Report summarizes the activities and accomplishments of the Land Information System (LIS) project during FY02. The overarching objective of the LIS project is to build a high-resolution, high-performance land surface modeling and data assimilation system to support a wide range of land surface research activities and applications.

This document has been prepared in accordance with the requirements of the Task Agreement GSFC-CT-2 under Cooperative Agreement Notice CAN-00-OES-01 Increasing Interoperability and Performance of Grand Challenge Applications in the Earth, Space, Life, and Microgravity Sciences, funded by NASA's ESTO Computational Technologies (formerly High Performance Computing and Communications) Project.

1.1 Project goals

As stated in the proposal and approved software engineering documents, the goals of LIS include:

- Realistic land surface modeling. LIS will simulate the global land surface variables using various land surface models, driven by atmospheric "forcing data" (e.g., precipitation, radiation, wind speed, temperature, humidity) from various sources.
- High performance computing. LIS will perform high-performance, parallel computing for near real-time, high-resolution land surface modeling research and operations.
- Efficient data management. The high-resolution land surface simulation will produce a huge data throughput, and LIS will retrieve, store, interpolate, re-project, sub-set, and backup the input and output data efficiently.
- Usability. LIS will provide intuitive web-based interfaces to users with varying levels of access to LIS data and system resources, and enforce user security policies.
- Interoperable and portable computing. LIS will incorporate the ALMA (Assistance for Land surface Modeling Activities) and ESMF (Earth System Modeling Framework) standards to facilitate inter-operation with other Earth system models. In order to demonstrate portability of LIS, the land surface modeling component will be implemented on a custom-designed Linux cluster and an SGI Origin 3000.

1.2 Land Surface Modeling Background

In general, land surface modeling seeks to predict the terrestrial water, energy and biogeochemical processes by solving the governing equations of the soil-vegetation-snowpack medium. Land surface data assimilation seeks to synthesize data and land surface models to improve our ability to predict and understand these processes. The ability to predict terrestrial water, energy and biogeochemical processes is critical for applications in weather and climate prediction, agricultural forecasting, water resources management, hazard mitigation and mobility assessment.

In order to predict water, energy and biogeochemical processes using (typically 1-D vertical) partial differential equations, land surface models require three types of inputs: 1) initial conditions, which describe the initial state of land surface; 2) boundary conditions, which describe both the upper (atmospheric) fluxes or states—also known as "forcings"—and the lower (soil) fluxes or states; and 3) parameters, which are a function of soil, vegetation, topography, etc., and are used to solve the governing equations.

The proposed LIS framework will include various components that facilitate global land surface modeling within a data assimilation system framework. The main software components of the system are:

- Land Data Assimilation System (LDAS): LDAS is a software system that integrates land surface models in a data assimilation framework.
- Land surface Models: LIS will include 3 different land surface models, namely, CLM, NOAH, and VIC.

These components are explained in detail in the LIS Software Design Document.

2 Approach

The overall approach for the LIS project involves parallel efforts throughout the three-year project directed towards the performance and interoperability goals of the original CAN, while adopting standard software engineering methodologies to ensure that the products are portable, reusable, extensible and maintainable.

Activities during FY02 have been focused on project initiation, infrastructure development, hires, and progress toward the FY02 milestones. Two kickoff meetings were held: a "local" meeting on February 13, 2002, and an "all investigator" meeting on March 4, 2002. We also initiated weekly local team meetings and monthly teleconferences following these meetings. The FY02 milestones and out-year milestones are given in the below, and are described in more detail on the LIS web site (http://lis.gsfc.nasa.gov).

2.1 FY02 Milestones

Project Start	16-JAN-2002
A) Software Engineering Plan	30-JUN-2002
E) Global LDAS Code Baselined	30-JUL-2002
B) FY02 Annual Report delivered	30-AUG-2002
f) Installation of Linux Cluster	30-AUG-2002

2.2 Out-year Milestones

H) Design policy for Interoperability and	28-FEB-2003

Community Delivery	
F) First Code Improvement	30-MAR-2003
I) Interoperability Prototype	30-JUL-2003
C) FY03 Annual Report	30-AUG-2003
G) Second Code Improvement	28-FEB-2003
J) Full Interoperability demonstrated	30-JUL-2004
K) Customer delivery Accomplished	30-AUG-2004
K) Present LIS to Review Board	30-AUG-2004
D) Final Report Delivered	28-FEB-2005

3 Scientific Accomplishments

The Land Information System is based largely on the Land Data Assimilation System (LDAS; http://ldas.gsfc.nasa.gov), and is the product of a 5-year effort towards the goal of modeling land surface states and fluxes, while relying as much as possible on observation-based parameter and forcing fields in order to avoid biases that are known to exist in forcing fields produced by atmospheric models. A recently submitted manuscript describes the science of LDAS in detail (Rodell et al., 2002).

4 Technology Accomplishments

As stated above, the first year of the LIS project has focused on project initiation, infrastructure development, hires, and progress toward the FY02 milestones. Two kickoff meetings were held

4.1 Design and Construction of the LIS Web Site

The LIS web site has been developed at the URL http://lis.gsfc.nasa.gov for communication and dissemination of the LIS documents and products. We have made extensive use of the web site in our milestone delivery, review and acceptance process as well as for communicating progress to our Review Board and the general public. For example, approved versions of all software engineering documents and source code are available to the public under "Milestones" as well as under "Documentation->Public". In addition, the members of the Review Board and the LIS team are available under "Contacts". In the future, the LIS web site will also serve as an interface to LIS models and data.

4.2 Development of the Software Engineering Plan

The LIS software engineering plan was submitted and approved in June, 2002. The software engineering plan includes components on our technical, management and quality assurance approaches for the LIS.

4.3 Requirements Gathering and System Design

A formal requirements gathering process was initiated in Spring, 2002, and included members of the LIS and LDAS teams as well as collaborators from Princeton, COLA, and NCEP. Requirements were gathered using (1) Internal meetings; (2) Joint meetings with the LIS and LDAS team; and (3) The requirements gathered during this process were included in the requirements document submitted in July, 2002, and formed the basis for the Software Design and Traceability Matrix also submitted in July 2002. The culmination of the system design is reflected in Figure 1 below, which illustrates the data flow, interfaces, standards, parallelization and user interface aspects of LIS.

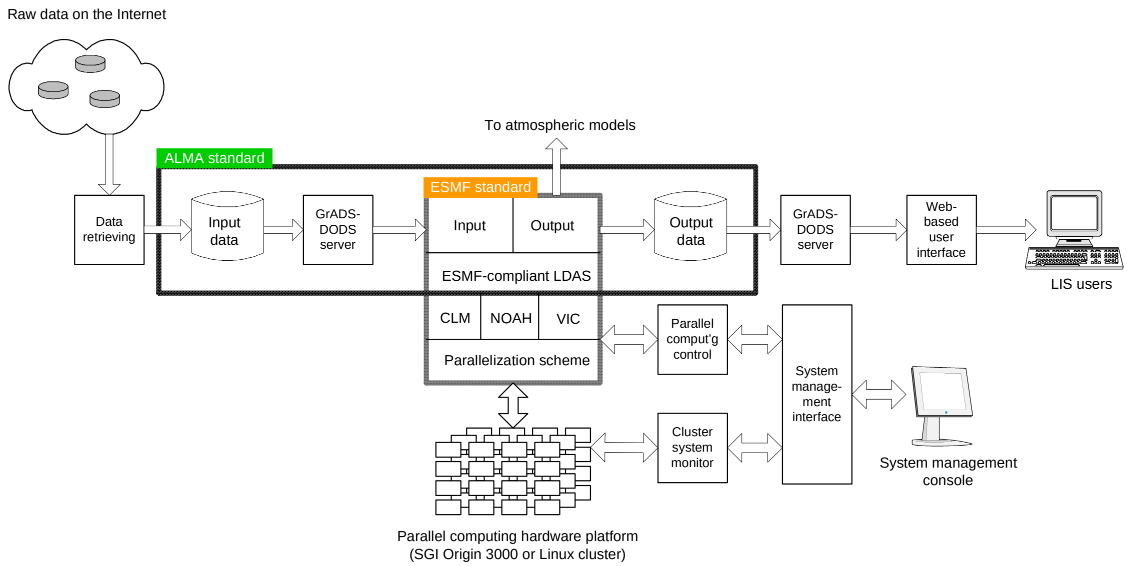

Figure 1: Overview of LIS software architecture and its components designed for the LIS cluster. A subset of the components, the LDAS and parallel computing implementation, will also be tested on SGI Origin platforms.

4.4 Baselining, Documentation, and Delivery of the LDAS Code

In addition to the substantial software engineering and design efforts described above, a parallel effort towards baselining, documentation and delivery of the LDAS code was accomplished for the July 2002 milestone. As given in our metrics section of our code documentation, the total number of lines of LDAS code is 31163, of which about half are executable. The baselining was conducted on the SGI Origin 3000 system at Ames, and the outcome of this effort suggests that certain routines LDAS will be straightforward to parallelize in order to meet our performance objectives. However, the analysis also suggests that substantial effort will also need to be devoted to data staging and I/O.

4.5 Design and Construction of the LIS Cluster

The LIS project decided to pursue the optional Linux cluster milestone based on the projected resource requirements of the LIS at 1 km spatial resolution. The LIS cluster has been developed as of August, 2002, and the physical architecture of the LIS Linux cluster is shown in Figure 2. The cluster consists of 192 computing nodes and 8 IO (input – output) nodes interconnected with 8 Ethernet switches.

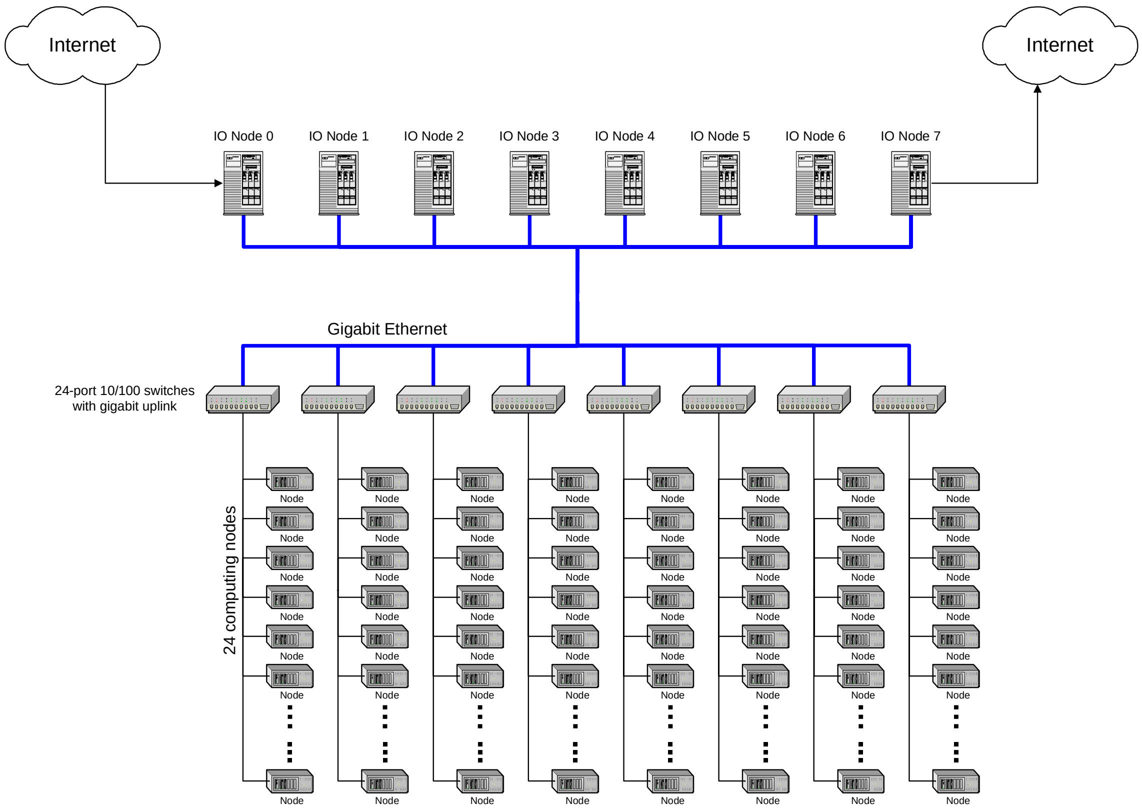

Figure 2: The physical architecture of the LIS Linux cluster. The cluster has 8 IO nodes and 192 compute nodes. Each IO node has dual Athlon CPUs, 2GB RAM and Gigabit NICs, and each compute node has a single Athlon CPU, 512MB RAM and a Fast Ethernet NIC.

5 Status/Plans

In addition to the Technological Accomplishments described above, the LIS team has made significant progress towards future milestones in the areas of interfacing with the ESMF, GrADS/DODS Server (GDS) development, and VIC code modification and documentation. Key progress in these areas is described below, followed by a discussion of plans for FY03.

5.1 Interfacing with the developing ESMF

In an effort to ensure that the developing ESMF meets the needs of land surface modeling, and will provide capabilities that can be successfully demonstrated in LIS, we have initiated several linkages with the ESMF team. Arlindo da Silva has been appointed as a member of our review board, and Paul Houser has been appointed as a member of the ESMF review board. Further, Christa Peters-Lidard attended and presented information on the LIS project at an ESMF team meeting held at GSFC in May, 2002.

Christa Peters-Lidard also attended the ESMF requirements meeting held in Washington, DC in May, 2002. Both Paul Houser and Christa Peters-Lidard have been invited to attend the upcoming ESMF meeting in September, and we plan to continue attending all of the open forum ESMF meetings.

5.2 Verification of GDS Capability to Handle Global 1-Degree Data

The ability of the GDS and GrADS to handle large data sets was verified by COLA. A 1 km global grid (University of Maryland land cover data; nominally 1/120 degree resolution on a rectangular latitude-longitude grid, which is approximately 930 m at the equator and 16 m at 89N) was loaded on a GDS at COLA, and was fully addressable, downloadable and displayable using a GrADS client. Actual data accessibility will depend on the server and client platforms and their memory limits. One needs to be able to allocate *approximately 1GB* of memory for a single global variable at 1 km resolution. If memory limitations on either the server and client machines prevent the full grid from being loaded as a single "subset" request, requests of this size could easily be made using client-side GrADS scripts that load the grid as several smaller subsets.

5.3 GDS Throughput Tests

A preliminary test of GDS performance under the worst-case scenario for delivering forcing data was measured at COLA. The worst case involves all 192 compute nodes requesting forcing data simultaneously from the 8 I/O nodes (24 compute nodes per I/O node). The machine used for the test was a single processor, Pentium III 800Mhz. This represents approximately 1/4 of the computing capacity of an LIS I/O node (dual processor, Athlon 1.7GHZ).

A GDS was installed and configured with a small binary-format data set. Load generator software (Apache JMeter) was used to simulate 24 compute nodes making simultaneous requests on the GDS. The request pattern was designed to mimic the expected pattern of forcing data requests - each simulated "node" made five data requests, ranging from 300 bytes to 1200 bytes of data each.

The GDS took an average of 800 ms to complete 24 simultaneous data requests. Thus the time it would take to initialize 24 nodes with a complete set of forcing data is 5*800ms or 4 seconds. This shows that even using a substantially slower platform than will be deployed for the LIS, and without any special performance optimizations, the GDS can achieve satisfactory performance under worst-case conditions.

5.4 Documentation of the VIC model

The VIC model code is written in ANSI C. The model code is available online and the documentation on the web is currently being maintained by the Land Surface Hydrology Group at the University of Washington (http://www.hydro.washington.edu/Lettenmaier/Models/VIC/VIChome.html). The user documentation provided at this website does not match with the latest version of the code

(Version 4.0.3). The Princeton modeling team is updating this documentation to be consistent with the online model documentation for the other land surface schemes as delivered in the July, 2002 milestone.

5.5 FY03 Plans

As shown above, the key FY03 milestones relate to interoperability and performance. Towards these milestones, we will be developing a rapid prototype LDAS parallelization, in consultation with the CT staff scientists, in order to assess the relative performance gains possible from parallel computation vs. parallel I/O. In addition, we will be developing several interface prototypes, including web configuration and visualization interfaces, GDS-LDAS interface, and LDAS-land surface model interfaces. We will be working closely with our collaborators at COLA, Princeton and NCEP, in addition to the CT software engineering team and the ESMF team to ensure that these interface designs are consistent with developing standards for Earth system modeling, and specifically, land surface and hydrologic modeling.

Further testing of GDS performance is planned for the near future, using actual example input and output data on a system with specifications close to those of the LIS I/O nodes. Support for ALMA/NetCDF-GDT convention data is being incorporated into GrADS and GDS. The COLA team will investigate the ESMF map projection/reprojection requirements and consider feasibility of incorporating those standards into GDS.

6 Points of Contact

Dr. Paul Houser, Co-PI
Head, Hydrological Sciences Branch
NASA-GSFC Code 974
Greenbelt, Maryland 20771
Office: 301/614-5772 Home: 410/872-0549
Fax: 301/614-5808 Cellular: 301/613-3782
E-mail: houser@hsb.gsfc.nasa.gov
Web: http://hsb.gsfc.nasa.gov

Dr. Christa Peters-Lidard, Co-PI and Project Manager
Research Scientist, Hydrological Sciences Branch
NASA-GSFC Code 974
Greenbelt, Maryland 20771
Office:301/614-5811 Fax:301/614-5808
E-mail: cpeters@hsb.gsfc.nasa.gov
Web: http://hsb.gsfc.nasa.gov/cpeters

7 List of Publications

None at this time.

8 List of Conference Presentations

None at this time.

9 List of Media References

None at this time.

10 List of Patents Filed

None at this time.

11 List of Students/Post Docs Trained

The LIS project enlisted the help of two undergraduate students this summer to assist with construction of the cluster. In addition, the project hired two post-docs through GEST for the duration of the three-year project.

11.1 Students Trained

- Uttam Majumder, SIECA Summer Intern, 2002
- Nikkia Anderson, Summer Intern, 2002

11.2 Post Docs Trained

- Sujay Kumar, Ph.D.
- Yudong Tian, Ph.D.

12 References

ALMA: http://www.lmd.jussieu.fr/ALMA/

Atlas, R. M., and R. Lucchesi, File Specification for GEOS-DAS Gridded Output. Available online at : http://dao.gsfc.nasa.gov/DAO_docs/File_Spec_v4.3.html , 2000.

Chen, F., K. Mitchell, J. Schaake, Y. Xue, H. Pan, V. Koren, Y. Duan, M. Ek, and A. Betts, "Modeling of land-surface evaporation by four schemes and comparison with FIFE observations", *J. Geophys. Res.*, 101, D3, 7251-7268, 1996.

CLM: http://www.cgd.ucar.edu/tss/clm/

Collatz G. J., C. Grivet, J. T. Ball, and J. A. Berry, J. A. "Physiological and Environmental Regulation of Stomatal Conductance: Photosynthesis and Transpiration: A Model that includes a Laminar Boundary Layer", *Agric. For. Meteorol.* , 5, pp 107 -- 136, 1991.

Derber, J. C., D. F. Parrish, and S. J. Lord, "The new global operational analysis system at the National Meteorological Center", *Wea. And Forecasting*, 6, pp 538-547, 1991.

ESMF: http://www.esmf.ucar.edu/
GrADS-DODS server: http://grads.iges.org/grads/gds/
Hamill, T. M., R. P. d'Entremont, and J. T. Bunting, "A description of the Air Force real-time nephanalysis model", *Wea. Forecasting,* 7, pp 238-306, 1992.
Hofstee, H. P., J. J. Likkien, and J. L. A. Van De Snepscheut "A Distributed Implementation of a Task Pool". *Research Directions in High-Level Parallel Programming Languages*, pp 338--348, 1991.
Jarvis, P. G., " The interpretation of leaf water potential and stomatal conductance found in canopies in the field", *Phil. Trans. R. Soc.* London, Ser. B, 273, pp 593 – 610, 1976.
Kopp, T. J. and R. B. Kiess, "The Air Force Global Weather Central cloud analysis model", *AMS 15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, pp 220-222, 1996.
LDAS: http://ldas.gsfc.nasa.gov/
NOAH: http://www.emc.ncep.noaa.gov/mmb/gcp/noahlsm/README_2.2.htm
Pfaendtner, J., S. Bloom, D. Lamich, M. Seablom, M. Sienkiewicz, J. Stobbie, and A. da Silva, "Documentation of the Goddard Earth Observing System (GEOS) Data Assimilation System – Version 1", *NASA Technical Memorandum* 104606, 4, pp 44, 1995.
Reynolds, C. A., T. J. Jackson, and W. J. Rawls, "Estimating available water content by linking the FAO Soil Map of the World with global soil profile databases and pedo-transfer functions" *American Geophysical Union, Fall Meeting, Eos Trans. AGU*, 80, 1999.
Richards, L. A., "Capillary conduction of liquids in porous media", *Physics*, 1, pp 318—333, 1931.
Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, and K. Mitchell, The Global Land Data Assimilation System. Submitted to Bull. Amer. Meteor. Soc., 2002.
Rogers, E., T. L. Black, D. G. Deaven, G. J. DiMego, Q. Zhao, M. Baldwin, N. W. Junker, and Y. Lin, "Changes to the operational "early" eta analysis / forecast system at the National Centers for Environmental Prediction" *Wea. Forecasting*, 11, pp 391-413, 1996.
Shapiro, R. "A simple model for the calculation of the flux of direct and diffuse solar radiation through the atmosphere", AFGL-TR-87-0200, Air Force Geophysics Lab, Hanscom AFB, MA.
Turk, F. J., G. Rohaly, J. D. Hawkins, E. A. Smith, A. Grose, F. S. Marzano, A. Mugnai, and V. Levizzani, "Analysis and assimilation of rainfall from blended SSM/I, TRMM, and geostationary satellite data", *AMS 10th Conf. On Sat. Meteor. and Ocean.*, Long Beach, CA, 9-14 January, pp 66-69, 2000.
VIC: http://www.hydro.washington.edu/Lettenmaier/Models/VIC/VIChome.html